

Laws of thermodynamics beyond the von Neumann regime

Dario Egloff,¹ Oscar C. O. Dahlsten,^{2,3,1} Renato Renner,¹ and Vlatko Vedral^{2,3}

¹*Institute for Theoretical Physics, ETH Zürich, 8093 Zurich, Switzerland*

²*Center for Quantum Technologies, National University of Singapore, Republic of Singapore*

³*Atomic and Laser Physics, Clarendon Laboratory,
University of Oxford, Parks Road, Oxford OX13PU, United Kingdom*

(Dated: July 3, 2012)

Standard thermodynamics is associated with the von Neumann entropy. There has recently been great interest in making thermodynamics more suitable for the quantum and nano-regime by taking inspiration from single-shot information theory, which replaces the von Neumann entropy with the more general smooth entropies. These reduce to the von Neumann entropy in the regime of asymptotically many identical and uncorrelated systems, the von Neumann regime. We propose generalisations of the laws of thermodynamics which analogously reduce to the standard laws in that same limit. We motivate these by analysing a very generic Szilard-engine model. We derive how much work one can extract in a single extraction as a function of the success-probability. Neither the initial nor final state is required to be thermal. In the appropriate limits we recover existing results.

General introduction.— There has long been a productive cross-fertilization between information theory and statistical mechanics. For example, Jayne’s maximum entropy approach to statistical mechanics was heavily inspired by Shannon information theory [1]. In the other direction, von Neumann’s quantum entropy was originally motivated through a thermodynamical argument, and is now the canonical measure of information in quantum information theory [2].

A recent development in information theory is the generalisation of the standard Shannon information theory, classical as well as quantum, to something tentatively called *smooth entropy* information theory. To our knowledge this first appeared in [3, 4]. The generalisation turned out to be necessary in the context of quantum cryptography. The problem with the standard theory is that the Shannon entropy (von Neumann entropy in the quantum case) only has the operational interpretations desired in a particular limit of extremely large numbers (asymptotic) of identical independently distributed states (i.i.d.). In cryptography one cannot allow oneself to make strong assumptions like this. In contrast the entropy(ies) used in the smooth entropy approach have the desired operational meanings for *all* states, including finite sized, correlated states and for single realizations of protocols, not just on average—thence the alternative name for the approach: *single-shot information theory*. In the appropriate limit the smooth entropy approach reduces to standard Shannon approach.

Given the history of fruitful cross-fertilization between the fields it is natural to consider whether this new approach to information theory is useful in statistical mechanics. To our knowledge this was first considered in [5]. The focus was on the relationship between information, quantified by entropy, and *work*. This relationship has been the centre of much intriguing, and arguably very productive debate, c.f. Maxwell’s daemon [6–8], Szilard’s engine [9], Landauer’s erasure [10], and Bennett’s reversible measurements [11]. Particularly important for

[5] and our considerations here is the notion of a Szilard engine [9] and Bennett’s extensions thereof [11]. It was shown in [5] that one can use smooth entropy to quantify the extractable work in a Szilard engine, making the expressions much more generally applicable than the corresponding Shannon entropy expression. This has been followed by several further results. In [12] it was shown how to interpret negative conditional entropy in these settings ([5] does not deal with conditional entropy). Very recently, in [13] and independently [14] the non-conditional case is considered in a significantly more sophisticated and general manner than [5]. These articles taken together give much hope that a neat and greatly generalised statistical mechanics, tentatively dubbed *single-shot statistical mechanics*, is emerging. A key advantage with this approach is that one can answer questions such as “how much work can I extract in any given go (single shot extraction) with a probability x of success”. In standard thermodynamics one is only concerned with averages.

In this Letter we take this much further. We give an expression for the extractable work in a single extraction which reduces to the expressions of [5, 13, 14] in the appropriate limits. We take the system constituting the working medium of the Szilard engine to have a given but arbitrary set of energy levels. Moreover it has an arbitrary probability distribution over these levels, representing our knowledge thereof. This corresponds to the eigenvalues of a density matrix ρ taken to be diagonal in the energy basis. Similarly we take the final energy and probability distributions to also be fixed but arbitrary. (In [13, 14] this is taken to be a thermal state on the same energy levels.) Our finding is that the maximal work that can be extracted given these initial and final conditions is determined by an apparently novel measure of how much a distribution majorizes another. We arrive at the result partly by combining single-shot concepts with techniques from [15–18] which have to our knowledge not previously been applied to this setting. We then use our result to generalise the laws of thermodynamics,

with particular emphasis on the second law.

We proceed as follows. We firstly briefly review some relevant definitions and existing results relating to single-shot statistical mechanics. We then define the setting we consider here, the work extraction game. We proceed to build up to our main statement by defining the novel majorisation measure in question, as well as an important tool we will adapt from [15, 16, 18]. We then give the main statement: how much work can one at most extract in the game given the initial and final conditions. A sketch of the proof is subsequently provided, as well as the actual proof which is in the technical supplement. Finally we discuss the differences to previous results and what the result tells us about the structure of single-shot thermodynamics.

Single-shot statistical mechanics, relevant key results.—We now briefly review key results that we shall later recover as special cases of our expression. (This is thus not an exhaustive list of all previous results). More specifically, the details of the models of work extraction in the different papers are not a priori identical, but we shall recover the same expressions for the extractable work within the model here.

In [5] an expression for the extractable work given an n -cylinder (part) Szilard engine was given as

$$W^\varepsilon = (n - H_{\max}^\varepsilon) kT \ln 2. \quad (1)$$

Here W^ε is the work that can be extracted for certain except with probability ε . H_{\max}^ε is the *smooth max entropy* of the density matrix representing the agent's initial knowledge about the state of the working medium. This is defined as $H_{\max}^\varepsilon(\rho) = \log(\text{Supp}^\varepsilon(\rho))$, with $\text{Supp}^\varepsilon(\rho)$ the support of ρ minimised over all states within ε trace distance of ρ . (Actually there is an alternative definition as well but they are both known to coincide up to an additive $\log \frac{1}{\varepsilon}$ term, so for simplicity we mention only one definition here). T is the temperature of the heat bath, and k Boltzmann's constant.

The more recent papers [13, 14] use work extraction models that are not a priori equivalent to each other, but the result for the extractable work recovered is the same in both papers. A key result obtained independently in both models is that given an initial state ρ and a final thermal state ρ_T over the same energy levels, the work that can be extracted with certainty up to ε failure probability is:

$$W^\varepsilon = kT \ln(2) D_0^\varepsilon(\rho || \rho_T), \quad (2)$$

where ρ is taken to be diagonal, ρ_T the corresponding thermal state on the same energy levels, and $D_0^\varepsilon(\rho || \rho_T)$ is the ε -smooth relative entropy of order 0 (see [19]). This reduces to $W = kT \ln 2 D(\rho || \rho_T)$ for the standard relative entropy in what we shall here call the *von Neumann regime*, where $\rho = \tau^{\otimes n}$, $n \rightarrow \infty$ and $\varepsilon \rightarrow 0$. That latter expression is well-established, see e.g. [20]. Moreover Eq. 2 reduces to Eq. 1 in the case of degenerate energy levels, as shown in [13].

The work extraction game.—The game has simple but minimal rules. (It will nevertheless not be trivial to analyse as there is a multitude of different strategies one may choose). There are three systems and a work-extraction agent. One system is the working medium, another is a heat bath of temperature T , and the last is the work reservoir. The agent wishes to transfer as much energy as possible into the work reservoir in a single extraction. We shall be concerned with quantifying how much energy can be transferred with certainty up to probability ε , calling this the work, W^ε .

The initial energy spectrum $\{E\}$ of the working medium is arbitrary. The initial density matrix ρ of the system is diagonal in the energy basis. The final energy spectrum $\{F\}$ and diagonal density matrix σ are also arbitrary. Whilst the initial and final energy spectra are arbitrary it should be noted that W^ε will turn out to depend on them, so they must be specified in order to calculate W^ε . The agent has a few elementary processes it can combine in any way it chooses: (i) it may couple the working medium to the heat bath. This has the effect of changing the probabilities (not the energy levels) in such a way that they approach, by some amount, the Gibbs thermal state for the given energy spectrum (wherein $p(E_i) = \exp[-E_i/kT]/Z$); (ii) it may change the energy levels (without altering the probabilities) by external intervention taking $\{E\}_j$ to $\{E\}_{j+1}$ where j labels the time step. Here the energy must be accounted for by being taken from or given to the work reservoir. In a given realisation the system is in one, possibly unknown, energy eigenstate and only changes to that particular eigenstate cost or yield work; (iii) It may permute, or equivalently relabel, the eigenstates. The combination of these elementary processes the agent chooses is called its *strategy*. Energy conservation is assumed throughout, in particular any energy leaving or entering the system must enter or leave the heat bath and/or the work-extraction system.

Relative mixedness.—We now introduce a measure of how much more mixed one state ρ is than another, σ , calling this the relative mixedness $M(\rho || \sigma)$. The definition of $M(\rho || \sigma)$ will later be justified by operational statements we will make.

Definition 1 (Relative mixedness). *The relative mixedness of two states ρ and σ with compact support and spectra $f(x_1)$, $g(x_2)$ respectively is given by*

$$M(\rho || \sigma) := \max m : \int_0^l f(x_1) dx_1 \geq \int_0^{lm} g(x_2) dx_2 \forall l, \quad (3)$$

For states with discrete spectra $\{\lambda_i\}$ one evaluates M for the associated step-function where the i -th 'block' of constant height has height λ_i and all blocks have width 1.

The logarithmic relative mixedness $\ln M(\rho || \sigma)$ (the \ln will turn out to make the interpretation neater) can be viewed as a measure of how much the spectrum of ρ majorises that of σ . If m is set to 1 the expression optimised

over in the definition in M reduces to the definition for whether (the spectrum of) ρ majorises σ , $\rho \succ \sigma$. We shall discuss more relations between M and majorisation later in the Letter.

Gibbs-rescaling.— We now describe a powerful insight from [15–17] which we employ here to bridge a particular gap between information theory and statistical mechanics: the fact that the former does not care about energy. In information theory, the Shannon/von Neumann entropy of a state, $-\sum_i p_i \log p_i$ is independent of the energies of the states involved. As the extractable work is expected to depend on the energy levels involved it follows that it is not expected to be uniquely determined by an entropy (or any other quantity that is independent of energy). A key way in which energy enters into statistical mechanics is that in a Gibbs state the probability of any given energy eigenstate state with energy E is given by $p_T(E) = \frac{\exp(-\frac{E}{kT})}{Z}$, where Z is the partition function. The insight we adapt from [15–17] is that we can take this bias into account by what essentially amounts to rescaling the density matrix's eigenvalue distribution by $p_T(E)$. After this rescaling the occupation probabilities will turn out to uniquely determine our expression for the extractable work. More specifically, we shall be employing an operation we term *Gibbs-rescaling* to the eigenvalue spectrum. Firstly we transform the spectrum $\{\lambda_i\}$ into the associated step-function. Then we take each block, rescale its height as $\lambda_i \mapsto \frac{\lambda_i}{e^{-\frac{E_i}{kT}}}$, and

its width $l = 1 \mapsto e^{-\frac{E_i}{kT}}$ such that the area of the new block is λ_i as before. We write this operation applied to a density matrix ρ as $G^T(\rho)$.

The main theorem.— Having defined the relative mixedness $M(\cdot||\cdot)$ and Gibbs-rescaling $G^T(\cdot)$ we are now ready to give the main statement.

Theorem 1. *In the work extraction game defined above, if one is given an initial density matrix $\rho = \sum_i \lambda_i |e_i\rangle\langle e_i|$ and final density matrix $\sigma = \sum_j \nu_j |f_j\rangle\langle f_j|$ with $\{|e_i\rangle\}, \{|f_j\rangle\}$ the respective energy eigenstates and both ρ and σ having finite rank, then the work W^ε one can extract with certainty except with ε probability respects*

$$W^\varepsilon \leq kT \ln \left(M \left(\frac{G^T(\rho)}{1-\varepsilon} || G^T(\sigma) \right) \right).$$

The proof of the statement is in the technical supplement. A rough intuition is that the Gibbs-rescaling is needed to take into account the bias imposed on the energy levels by the Gibbs statistics, and up to that bias, only the amount of majorisation matters as the work extraction process lowers the majorisation amount, the relative mixedness.

Figure 1 gives an example of a simple application of the theorem.

We now discuss the implications of the statement and develop the argumentation further. Firstly we consider how to generalise the laws of thermodynamics for them to be defined and correct in our model.

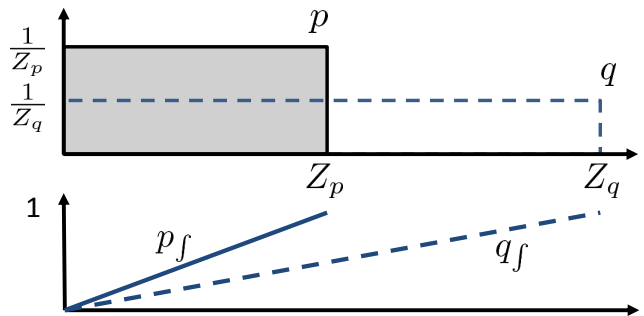


Figure 1: Case of initial and final states being Gibbs states. The Gibbs-rescaling takes a Gibbs state with partition function $Z_p(Z_q)$ to a uniform distribution $p(q)$ of width $Z_p(Z_q)$ and height $\frac{1}{Z_p}$ ($\frac{1}{Z_q}$) (upper graph). The integrals p_f and q_f (lower graph) are used to evaluate the relative mixedness. Consider firstly $\varepsilon = 0$. m defined in the relative mixedness definition must satisfy $p_f(ml) \geq q_f(l) \forall l$, and we see this holds for $m \leq \frac{Z_q}{Z_p}$, implying that $W^0 \leq kT \ln \frac{Z_q}{Z_p}$. For the case of $\varepsilon \neq 0$ p_f is replaced with $\frac{p_f}{1-\varepsilon}$ and one sees $m \leq \frac{Z_q}{Z_p(1-\varepsilon)}$. Thus in this case $W^\varepsilon \leq kT \ln \frac{Z_q}{Z_p} + kT \ln \frac{1}{1-\varepsilon}$. A special case of the above corresponds to a generalisation of Landauer's erasure principle, as we can take the initial state to be two thermalised degenerate levels at energy 0 (i.e. one unknown bit) and the final state to be thermalised but with one energy level at 0 and the other extremely high (i.e. a bit taking one value only). Then we see $W^\varepsilon \leq -kT \ln 2 + kT \ln \frac{1}{1-\varepsilon}$, so that work would have to be invested if ε is suitably small. (The reverse direction is also possible, corresponding to a single-qubit Szilard engine).

0th and 1st laws of thermodynamics.— The 0th law can be stated as: *There exists for every thermodynamic system in equilibrium a property called temperature. Equality of temperature is a necessary and sufficient condition for thermal equilibrium.* This also holds after our generalisation. In particular we are still assuming heat baths that take the working medium closer to a Gibbs thermal state upon interaction.

The first law however is more subtle as it involves distinguishing between heat and work and we do generalise the definition of work and accordingly that of heat. The law is normally stated as $dU = dQ - dW$ where $U = \text{tr}(\rho H)$ is the internal energy of the working medium with Hamiltonian H , Q is ‘heat’ and W ‘work’. The essential idea behind the splitting of the internal energy into two terms in such a manner is that one imagines three systems: a heat bath, a working medium, and an energy storing system (e.g. a weight that is lifted or an atom that is excited). Energy transfers to/from the working medium from the heat bath are called dQ and those to/from the storage system dW . Our model preserves this distinction of three systems and by assumption energy conservation of energy holds in any given extraction, i.e. $dE_{\text{sys}} = -dE_{\text{bath}} - dE_{\text{reservoir}}$. However we are going beyond talking about the *average* internal en-

ergy of the working medium and make statements about how this energy changes in individual extractions. Our generalisation of the first law can be stated as

In any given extraction, with $p \geq 1 - \varepsilon$

$$dE_{sys} = -dE_{bath} - dW^\varepsilon - dE_{rest} \equiv dQ - dW^\varepsilon. \quad (4)$$

Note that we are breaking the energy change of the work reservoir into two parts, dW^ε and dE_{rest} . The idea behind this is that only predicted energy transfer should count as work. One may for example imagine buckets lifting water out of a mine up to a certain height (or as a quantum example an atom excited up to a certain level). The height at which the buckets are tipped into a reservoir is specified in advance. If they go higher than this, the extra potential energy will be transferred to other degrees of freedom associated with the reservoir system, e.g. into movement of the water. We call this extra dE_{rest} energy wasted and group it with the generalised heat $-dQ = dE_{bath} + dE_{rest}$. This second term is not present in the von Neumann regime (asymptotic i.i.d. with vanishing ε) as there the work extraction is deterministic.

Second law.—Consider the so-called Kelvin statement of the second law:

No process is possible in which the sole result is the absorption of heat from a reservoir and its complete conversion into work.

This law, or more specifically a natural quantitative generalisation of this law, holds also in our more general setting. We show in the appendix that for given states of the working medium A and B respectively, $W^\varepsilon(A \rightarrow B) + W^\varepsilon(B \rightarrow A) \leq W^{2\varepsilon}(A \rightarrow A)$. We call this the ‘triangle inequality’. It implies the following generalisation of the second law.

$$\sum_{i=0}^{m-1} W^\varepsilon(A_i \rightarrow A_{i+1}) \leq W^{m\varepsilon}(A \rightarrow A) \text{ if } A_m = A_1. \quad (5)$$

If $\varepsilon = 0$, which would be the case if one wants the system to be reset at the end with certainty, then the right-hand-side of the inequality is 0. Note that one may still get work out in a single cycle by letting ε depend on i . One could e.g. take a larger risk of failure when resetting compared with extracting and then get net work out from a cycle if successful.

The second law is also closely related to entropy increasing with time and one may wonder what the corresponding generalisation of the statement is. A particular standard expression is that

$$\Delta(S - \beta\langle E \rangle) \geq 0 \quad (6)$$

where S and $\langle E \rangle$ are the von Neumann entropy and expected energy of a system interacting with a heat-bath with inverse temperature β . (Δ indicates the change in these values during the interaction.) This actually still holds in our more general model; we show this in the

technical supplement. However, crucially, it is not *sufficient* to guarantee that an evolution is possible. Instead it should be replaced by the statement that a state change $\rho \rightarrow \rho'$ due to a thermalisation with a heat-bath at temperature T is possible if and only if

$$\ln(M(G^T(\rho)||G^T(\rho'))) \geq 0. \quad (7)$$

This is a significant strengthening of the restriction on second law-type entropy increase. There are processes that respect Eq. 6 but violate Eq. 7. A simple example is to consider degenerate energy levels so that $\Delta\langle E \rangle = 0$, and take three levels with probabilities $(1/2 \ 1/2 \ 0)^T \rightarrow (2/3 \ 1/6 \ 1/6)^T$. Then $\Delta S \approx 0.25$ but $\ln(M(G^T(\rho)||G^T(\rho')))$ is negative so this evolution is not possible according to our model. The inequivalence of entropy and majorisation displayed here has been previously noted in the context of the second law [15, 16]. The reason this has not received more attention to date is presumably that in the von Neumann regime this inequivalence disappears. More precisely, if we take a tensor product of n identical states each with von Neumann entropy S and let $n \rightarrow \infty$, then with asymptotically small error we may approximate the distribution over states as a uniform distribution with value 2^{-nS} in the range $[0, 2^{-nS}]$. Increasing S now gives a flatter and wider uniform distribution, majorised by the initial distribution.

The fact that Eq. 7 can be experimentally violated without violating the standard expression, implies an interesting potential test of our theory. More specifically, in our theoretical analysis Eq. 7 follows as a mathematical consequence, so the experimental question concerns whether our model is appropriate to describe physical systems interacting with their environment.

Recovering existing results.—Eq. 2 and accordingly 1 are special cases of our main result. Eq. 2 corresponds to the case where the final state ρ_T is demanded to have the same eigenspectrum and be a Gibbs state ($\rho_T = \sum p_T(E_i)|e_i\rangle\langle e_i|$). We show in the technical supplement that Eq. 2 is indeed recovered in that limit. (It was also shown in Fig. 1 that Landauer’s erasure principle is recovered in a simple manner.)

Outlook.—It is particularly important to test the results experimentally in non-equilibrium quantum systems, as outlined above. On the theoretical side one should also consider the maximal risk-taking type of strategies of [5] and the case of quantum side knowledge as in [12]. Finally we note the striking similarities between what is discussed above and the question of quantifying entanglement in the non-asymptotic regime. It was shown in a seminal paper by Nielsen [21] that majorisation is the central quantity there and we anticipate that many of the qualitative and quantitative results from our work can be applied also in that context.

Acknowledgements.—We acknowledge support from the National Research Foundation (Singapore), the Ministry of Education (Singapore), the Swiss National Science Foundation (grant No. 200020-135048) and the European Research Council (grant No. 258932). This

research was partly carried out in connection with DE's Master's thesis at ETH Zurich.

-
- [1] E. Jaynes, *The Physical Review* **106**, 620 (1957).
 - [2] J. Von Neumann, *Mathematical Foundations of Quantum Mechanics*, 1 ed. (Princeton University Press, 1955).
 - [3] R. Renner, *Security of Quantum Key Distribution*, PhD thesis, ETH Zürich, 2005, Also available as quant-ph/0512258.
 - [4] R. Renner and S. Wolf, Smooth Rényi entropy and applications, in *Proc. IEEE Int. Symp. Info. Th.*, p. 233, 2004.
 - [5] O. C. O. Dahlsten, R. Renner, E. Rieper, and V. Vedral, *New. J. Phys.* **13**, 053015 (2011).
 - [6] J. Maxwell, *Theory of Heat*, 4 ed. (Longmans, Green, London, 1985).
 - [7] H. Leff and A. F. Rex, *Maxwell's demon 2: Entropy, classical and quantum information, computing* (CRC Press, Abingdon, 2003).
 - [8] K. Maruyama, F. Nori, and V. Vedral, *Rev. Mod. Phys.* **81**, 1 (2009).
 - [9] L. Szilárd, *Z. Phys.* **53**, 840 (1929).
 - [10] R. Landauer, *Res. Dev* **5**, 183 (1961).
 - [11] C. H. Bennett, *Stud. Hist. Philos. Sci. B* **34**, 501 (2003).
 - [12] L. Rio, J. Aberg, R. Renner, O. Dahlsten, and V. Vedral, *Nature* **474**, 61 (2011).
 - [13] J. Aberg, pre-print **arXiv:1110.6121** (2011).
 - [14] M. Horodecki and J. Oppenheim, pre-print **arXiv:1111.3834** (2011).
 - [15] E. Ruch and A. Mead, *Theor. Chim. Acta* **41**, 95 (1976).
 - [16] E. Ruch, *Theor. Chim. Acta* **38**, 167 (1975).
 - [17] A. Mead, *J. Chem. Phys.* **66**, 95 (1977).
 - [18] E. Ruch, R. Schraner, and T. Seligman, *J. Chem. Phys.* **69**, 386 (1978).
 - [19] N. Datta, *Information Theory, IEEE Transactions on* **55**, 2816 (2009).
 - [20] M. Donald, *Journal of Statistical Physics* **49** (1987).
 - [21] M. Nielsen, *Phys. Rev. Lett.* **83** (1999).
 - [22] G. H. Hardy, I. E. Littlewood, and G. Polya, *Inequalities* (CUP, 1952).

I. TECHNICAL SUPPLEMENT

We first give certain definitions and lemmas needed to prove the main theorem.

Definition 2 (Gibb's rescaling). Consider a density matrix $\rho = \sum_{i=1}^n \lambda_i |e_i\rangle\langle e_i|$ with eigenvalues $\{\lambda_i\}_{i=1}^n$ and take the energy eigenstates of the system to be $\{|e_i\rangle\}_{i=1}^n$ with energies $\{E_i\}_{i=1}^n$ respectively. There is an associated step function for the spectrum, $\lambda(xn) = \lambda_{\lceil xn \rceil}$ where $x \in (0, 1]$. Similarly there is an energy step function $E(xn) = E_{\lceil xn \rceil}$ where $x \in (0, 1]$. The Gibbs rescaling associated with temperature T combines $\lambda(xn)$ and $E(xn)$ to a new function $G^T(y)$ given by

$$G^T \left(\int_0^x e^{-\frac{E_{\lceil yn \rceil}}{kT}} dy \right) = \frac{\lambda_{\lceil xn \rceil}}{e^{-\frac{E_{\lceil xn \rceil}}{kT}}}.$$

It follows that $G^T(y)$ is defined on $(0, Z]$, with $Z = \sum_{j=1}^n \exp\left(-\frac{E_j}{kT}\right)$ the partition function. Moreover $G^T(y)$ is a probability distribution satisfying $\int_0^Z G^T(y) dy = 1$.

Definition 3 (Notation). $\vec{s} \in \{0, 1\}^m$: path variable for a game with m work extractions (subsequently called "steps"): $s_j = 1$: system is in chosen states for work extraction

$s_j = 0$: system is not in chosen states for work extraction
 \hat{s}_j is the complement of s_j : $s_j = 1 \Leftrightarrow \hat{s}_j = 0$ and $s_j = 0 \Leftrightarrow \hat{s}_j = 1$

$w_{\vec{s}}^j$: logarithmical work ($kT \ln(w_{\vec{s}}^j) = W_{\vec{s}}^j$) one extracted in the step j on the path \vec{s} .

w^j : The logarithmical work one tries to get out in the step j .

w : total logarithmical work demanded in order to call the total extraction successful.

$\eta_{\vec{s}}^j$: probability of doing step j on the path \vec{s} .

P_S : total probability of success: $P_S = \sum_{\vec{s} \in G} \prod_j \eta_{\vec{s}}^j$.

$\phi_{\vec{s}}^j$: state of the system after step j if the previous evolution of the system is given by the path \vec{s}

$\lambda_{\vec{s}}^j$: eigenvalues of state $\phi_{\vec{s}}^j$.

$p_{\vec{s}}^j = G(\phi_{\vec{s}}^j)$: Gibb's rescaled probability distribution after step j (before thermalizing) conditioned on the previous steps on path \vec{s} .

$p_{\vec{s},t}^j$: Gibb's rescaled probability distribution after step j (after thermalizing) conditioned on the previous steps on path \vec{s} .

q : final Gibb's rescaled probability distribution, conditioned on successful work extraction:

$$q = \sum_{\vec{s} \in G} \frac{\prod_j \eta_{\vec{s}}^j}{P_S} p_{\vec{s},t}^j$$

B_j : Bistochastic matrix one chooses after step j by thermalizing the system (this has to be the same for all paths).

$E^j(x)$: Energy of the level labeled by x after step j .

$\Theta_U(x)$: Step function on U :

$$\Theta_U(x) = \begin{cases} 1 & : \text{for } x \in U \\ 0 & : \text{else} \end{cases}$$

Definition 4 (Block). For $a < b$ the interval $(a, b]$ is said to be a block corresponding to a level k , if $p_{\vec{s}}^j$ is constant on this interval $\forall \vec{s}$.

Definition 5 (Thermalisation). If after step j one choses to do a thermalization one gets no change in the work reservoir and the Gibb's rescaled probability afterwards is given by:

$$p_{\vec{s},t}^j = B_j(p_{\vec{s}}^j)$$

where $B_j(\cdot)$ is associated with a bistochastic matrix B_j , acting on the vector $\vec{p}_k = p_{\vec{s}}^j(a_k)$, where $(a_{k-1}, a_k]$ with $a_{k-1} < a_k$ are blocks corresponding to levels k with $a_k - a_{k-1} = a_l - a_{l-1} \forall k, l$. $p_{\vec{s},t}^j$ is then defined by $p_{\vec{s},t}^j(x) = B_j \vec{p}_k$ for $x \in (a_{k-1}, a_k]$. B_j does not depend on the path \vec{s} .

Note that a permutation matrix is bistochastic so one may permute the levels in the thermalisation step. We use this to simplify the notation of the work extraction definition. We take the levels lowered or raised to form the first l levels.

Definition 6 (Work extraction). To do a work extraction in step j one first defines an interval $(0, a] = \bigcup_{k=1}^l (a_{k-1}, a_k]$, where $(a_{k-1}, a_k]$ correpond to levels $\{1, \dots, l\}$, on which one wants to change the energy by $\Delta E = -kT \ln(w_{\vec{s}|s_j=1}^j)$. The remaining levels remain untouched.

For $x \in (0, 1]$, the eigenvalues of the levels after step j , conditioned on the previous state are given by:

In the case $s_j = 1$ (state of the system is found to be in the levels corresponding to $(0, a]$):

$$\lambda_{\vec{s}|s_j=1}^j(\lceil xn \rceil) = \Theta_{(0,a]}(Z_{j-1}x) \frac{\lambda_{\vec{s}}^{j-1}(\lceil xn \rceil)}{\eta_{\vec{s}|s_j=1}^j}$$

where $\eta_{\vec{s}|s_j=1}^j = \int_0^a \lambda_{\vec{s}}^{j-1}\left(\left\lceil \frac{xn}{Z_{j-1}} \right\rceil\right) dx$.

In the case $s_j = 0$ (state of the system is not in the levels corresponding to $(0, a]$):

$$\lambda_{\vec{s}|s_j=0}^j(\lceil xn \rceil) = \Theta_{(a, Z_{j-1}]}(Z_{j-1}x) \frac{\lambda_{\vec{s}}^{j-1}(\lceil xn \rceil)}{\eta_{\vec{s}|s_j=0}^j}$$

where $\eta_{\vec{s}|s_j=0}^j = \int_a^{Z_{j-1}} \lambda_{\vec{s}}^{j-1}\left(\left\lceil \frac{xn}{Z_{j-1}} \right\rceil\right) dx$. In this case one has no change in the work reservoir: $w_{\vec{s}|s_j=0}^j = 1$.

As mentioned in the main body, our work extraction game is essentially to combine the elementary processes of thermalisation and work extraction in any sequence in order to extract as much work (or input as little as possible) whilst going from a defined initial state to a defined final state.

Definition 7 (Our work extraction game). *There are three systems and a work-extraction agent. One system is the working medium, another is a heat bath of temperature T , and the last is the work reservoir.*

The initial energy spectrum $\{E\}$ of the working medium is arbitrary but given. The initial density matrix ρ of the same is diagonal in the energy basis. The final energy spectrum $\{F\}$ and diagonal density matrix σ are also arbitrary but given.

The agent can combine thermalisation (defined above) and work extraction (also defined above) in any sequence. This sequence, together with the specifications for each step is called its strategy.

A given work extraction will transfer some energy ν to the work extraction reservoir. Before the extraction the agent must specify W . If $\nu \geq W$ the work extraction is termed successful. The probability of success is called $1 - \epsilon$.

We shall be interested in bounding W given ϵ and the initial and final conditions. We break the calculation into several lemmas which will later be combined to prove the main theorem.

The following lemma is relevant for quantifying how much flatter the Gibb's rescaled distribution is after a work-extraction.

Lemma 2. *For $x \in (0, Z_j]$ (with Z_j the partition function after step j), by doing a work extraction in step j , the Gibb's rescaled probability distribution after the step, conditioned on the previous steps on path \vec{s} are given by: In the case $s_j = 1$:*

$$p_{\vec{s}|s_j=1}^j(x) = \Theta_{(0, aw_{\vec{s}|s_j=1}^j]}(x) \frac{p_{\vec{s}}^{j-1} \left(\frac{x}{w_{\vec{s}|s_j=1}^j} + a \right)}{w_{\vec{s}|s_j=1}^j \eta_{\vec{s}|s_j=1}^j}$$

In the case $s_j = 0$:

$$p_{\vec{s}|s_j=0}^j(x) = \Theta_{(aw_{\vec{s}|s_j=0}^j, Z_j]}(x) \frac{p_{\vec{s}}^{j-1} (x - aw_{\vec{s}|s_j=0}^j + a)}{\eta_{\vec{s}|s_j=0}^j}$$

Proof. Case $s_j = 1$:

Let $w = w_{\vec{s}}^j$, $p^j = p_{\vec{s}}^j$, $p^{j-1} = p_{\vec{s}}^{j-1}$, $\lambda^j = \lambda_{\vec{s}}^j$ and $\eta^j = \eta_{\vec{s}}^j$. Let $x \in (0, aw] \cap (0, Z_j]$ and $b \in (0, \infty)$ such that

$$\int_0^b \exp \left(- \frac{E_{\lfloor \frac{yn}{w} \rfloor}^{j-1}}{kT} \right) dy = x.$$

$$\begin{aligned} p^j(x) &= p^j \left(\int_0^b \exp \left(- \frac{E_{\lfloor \frac{yn}{w} \rfloor}^{j-1}}{kT} \right) dy \right) \\ &= p^j \left(\int_0^b \exp \left(- \left(\frac{E_{\lfloor \frac{yn}{w} \rfloor}^{j-1}}{kT} - kT \ln(w) \right) \right) \frac{1}{w} dy \right) \\ &= p^j \left(\int_0^{b/w} \exp \left(- \left(\frac{E_{\lfloor zn \rfloor}^{j-1}}{kT} - kT \ln(w) \right) \right) dz \right) \\ &= \frac{\lambda_{\lfloor \frac{bn}{w} \rfloor}^j}{\exp \left(- \left(\frac{E_{\lfloor \frac{bn}{w} \rfloor}^{j-1}}{kT} - kT \ln(w) \right) \right)} \\ &= \frac{\lambda_{\lfloor \frac{bn}{w} \rfloor}^{j-1}}{\exp \left(- \left(\frac{E_{\lfloor \frac{bn}{w} \rfloor}^{j-1}}{kT} \right) \right)} w \eta^j \\ &= \frac{1}{w \eta^j} p^{j-1} \left(\int_0^{b/w} \exp \left(- \left(\frac{E_{\lfloor zn \rfloor}^{j-1}}{kT} \right) \right) dz \right) \\ &= \frac{1}{w \eta^j} p^{j-1} \left(\int_0^b \frac{\exp \left(- \left(\frac{E_{\lfloor \frac{yn}{w} \rfloor}^{j-1}}{kT} \right) \right)}{w} dy \right) \\ &= \frac{1}{w \eta^j} p^{j-1} \left(\frac{x}{w} \right) \end{aligned}$$

One easily sees that $p^j(x) = 0$ for $x \geq \frac{a}{w}$. The proof for the case $s_j = 0$ is analogous. \square

The next lemma is in contrast concerned with how much *wider* the Gibb's rescaled distribution is after the work extraction (recall that the distribution has support on $(0, Z]$ where Z is the partition function).

Lemma 3. *The partition function Z_j immediately after step j is given by:*

$$Z_j = Z_{j-1} + aw_1 - a$$

Proof. Let $(0, a]$ be an interval consisting of blocks corresponding to the levels $\{1, \dots, l\}$ and n such that it can

split the interval $(a, Z_j]$ into $n - l$ blocks.

$$\begin{aligned}
Z_j &= \sum_k e^{\frac{-E_k^j}{kT}} \\
&= \sum_{k=1}^l e^{\frac{-E_k^j}{kT}} + \sum_{k=l+1}^n e^{\frac{-E_k^j}{kT}} \\
&= \sum_{k=1}^l e^{\frac{-(E_k^{j-1} - kT \ln(w_1))}{kT}} + \sum_{k=l+1}^n e^{\frac{-E_k^j}{kT}} \\
&= w_1 \underbrace{\sum_{k=1}^l e^{\frac{-E_k^{j-1}}{kT}}}_a + \sum_{k=l+1}^n e^{\frac{-E_k^{j-1}}{kT}} \\
&= w_1 a - a + \sum_{k=1}^l e^{\frac{-E_k^{j-1}}{kT}} + \sum_{k=l+1}^n e^{\frac{-E_k^{j-1}}{kT}} \\
&= Z_{j-1} + aw_1 - a
\end{aligned}$$

out of which the lemma follows. \square

We now combine the two previous lemmas to gain another relation between the Gibb's rescaled distribution and steps j and $j - 1$. We shall use this later in an iterative manner to relate the very first and final Gibb's rescaled distributions.

Lemma 4. *The Gibb's rescaled probability distributions at steps j and $j - 1$ respectively satisfy the relation*

$$p_{\vec{s},t}^{j-1}(x) = \sum_{k=0,1} w_{\vec{s}|s_j=k}^j \eta_{\vec{s}|s_j=k}^j p_{\vec{s}|s_j=k}^j \left(x w_{\vec{s}|s_j=k}^j + c_{\vec{s}|s_j=k}^j \right)$$

with constants $c_{\vec{s}|s_j=1}^j = 0$ and $c_{\vec{s}|s_j=0}^j = aw^j - a$.

Proof. Let $w_k = w_{\vec{s}|s_j=k}^j$, $p_k^j = p_{\vec{s}|s_j=k}^j$, $p^{j-1} = p_{\vec{s},t}^{j-1}$, $\eta_k = \eta_{\vec{s}|s_j=k}^j$

Let $c_0 = aw_1 - a$ and $c_1 = 0$ and $c_{\vec{s}|s_j=k}^j = c_j$. Then:

$$\begin{aligned}
&\eta_0 w_0 p_0^j(xw_0 + c_0) + \eta_1 w_1 p_1^j(xw_1 + c_1) \\
&= \eta_0 p_0^j(x + aw_1 - a) + \eta_1 w_1 p_1^j(xw_1) \\
&= \Theta_{(aw_1, Z_j]}(x + aw_1 - a) p^{j-1}(x) + \Theta_{(0, aw_1]}(xw_1) p^{j-1}(x) \\
&= \Theta_{(a, Z_j - aw_1 + a]}(x) p^{j-1}(x) + \Theta_{(0, a]}(x) p^{j-1}(x) \\
&= \Theta_{(0, Z_{j-1}]}(x) p^{j-1}(x)
\end{aligned}$$

\square

Lemma 5 (Induction step). *Let $j \in \{1, \dots, m\}$. Let $l \in (0, Z_j]$. Let $\vec{s}' \in \{0, 1\}^{m-j-1}$. Define $\vec{s}_1 = (s_1, \dots, s_j, 1, s'_1, \dots, s'_{m-j-1})$ and $\vec{s}_0 = (s_1, \dots, s_j, 0, s'_1, \dots, s'_{m-j-1})$. Then:*

$$\begin{aligned}
&\sum_{\vec{s} \in \{0,1\}^j} \int_0^l \tau_t^j \circ p_{\vec{s}_0,t}^j(x) dx \\
&\geq \sum_{\vec{s} \in \{0,1\}^j} \int_0^l \left(w^{j+1} \eta_{\vec{s}_1}^{j+1} \tau_t^{j+1} \circ p_{\vec{s}_1,t}^{j+1}(xw^{j+1}) \right. \\
&\quad \left. + \eta_{\vec{s}_0}^{j+1} \tau_t^{j+1} \circ p_{\vec{s}_0,t}^{j+1}(x) \right) dx
\end{aligned}$$

where τ_t^j is the permutation, which maximizes the left hand side, while τ_t^{j+1} is the one which maximizes the right hand side.

Proof. Let $p_1 = p_{\vec{s}_1,t}^{j+1}$, $p_0 = p_{\vec{s}_0,t}^{j+1}$, $\eta_1 = \eta_{\vec{s}_1}^{j+1}$, $\eta_0 = \eta_{\vec{s}_0}^{j+1}$, $w = w^{j+1}$

$$\begin{aligned}
&\sum_{\vec{s} \in \{0,1\}^j} \int_0^l \tau_t^j \circ p_{\vec{s}_0,t}^j(x) dx \\
&= \sum_{\vec{s} \in \{0,1\}^j} \int_0^l \left(\eta_1 w \tau_t^j \circ p_1(xw) + \eta_0 \tau_t^j \circ p_0(x + aw - a) \right) dx \\
&= \sum_{\vec{s} \in \{0,1\}^j} \int_0^{l_1} \eta_1 w \tilde{\tau} \circ p_1(xw) \\
&\quad + \sum_{\vec{s} \in \{0,1\}^j} \int_a^{\tilde{l}+a-l_1} \eta_0 \tilde{\tau} \circ p_0(x + aw - a) dx
\end{aligned}$$

Where $\tilde{\tau}$ reorders $\sum_{\vec{s} \in \{0,1\}^j} p_1$ in descending order in $(0, aw]$ and $\sum_{\vec{s} \in \{0,1\}^j} p_0$ in $(aw, Z_j]$. This is possible since p_1 and p_0 have disjoint support, also for different \vec{s} , since a in definition 6 has to be chosen indepentantly of the path. (See lemma 2).

$l_1 \in (0, \min(a, l)]$ is a value which maximizes the right hand side of the last line. Using the same argument backwards after changing variables, we get:

$$\begin{aligned}
&\sum_{\vec{s} \in \{0,1\}^j} \int_0^l \tau_t^j \circ p_{\vec{s}_0,t}^j(x) dx \\
&= \sum_{\vec{s} \in \{0,1\}^j} \int_0^{l_1} \eta_1 w \tilde{\tau} \circ p_1(xw) + \sum_{\vec{s} \in \{0,1\}^j} \int_{aw}^{\tilde{l}+aw-l_1} \eta_0 \tilde{\tau} \circ p_0(x) dx \\
&= \sum_{\vec{s} \in \{0,1\}^j} \int_0^l \left(\eta_1 w \tau^{j+1} \circ p_1(xw) + \eta_0 \tau^{j+1} \circ p_0(x) \right) dx
\end{aligned}$$

Applying any bistochastic matrix B on the probabilities p_0 and p_1 and reordering in descending order with τ_t^{j+1}

afterwards, we get:

$$\begin{aligned}
& \sum_{\vec{s} \in \{0,1\}^j} \int_0^l \tau_t^j \circ p_{\vec{s}_0,t}^j(x) dx \\
& \geq \int_0^l \tau_t^{j+1} \circ B \circ (\tau^{j+1})^{-1} \circ \\
& \quad \sum_{\vec{s} \in \{0,1\}^j} (\eta_1 w \tau^{j+1} \circ p_1(xw) + \eta_0 \tau^{j+1} \circ p_0(x)) dx \\
& = \sum_{\vec{s} \in \{0,1\}^j} \int_0^l \left(w \eta_1 \tau_t^{j+1} \circ B \circ p_1(xw) \right. \\
& \quad \left. + \eta_0 \tau_t^{j+1} \circ B \circ p_0(x) \right) dx \\
& = \sum_{\vec{s} \in \{0,1\}^j} \int_0^l \left(w \eta_1 \tau_t^{j+1} \circ p_{\vec{s}_1,t}^{j+1}(xw) + \eta_0 \tau_t^{j+1} \circ p_{\vec{s}_0,t}^{j+1}(x) \right) dx
\end{aligned}$$

Where the inequality follows out of the inequality $Bp \succ p$ for any bistochastic matrix B and vector p , which is proved in [22]. \square

Theorem (Main, called Theorem 1 in main body). *In the work extraction game defined above, if one is given an initial density matrix $\rho = \sum_i \lambda_i |e_i\rangle\langle e_i|$ and final density matrix $\sigma = \sum_j \nu_j |f_j\rangle\langle f_j|$ with $\{|e_i\rangle\}$, $\{|f_j\rangle\}$ the respective energy eigenstates and both ρ and σ having finite rank, then the work W^ϵ one can extract with certainty except with ϵ probability respects*

$$W^\epsilon \leq kT \ln \left(M \left(\frac{G(\rho)}{1-\epsilon} || G(\sigma) \right) \right).$$

Proof. Define $p_{\vec{s}'}^0 = p$. W.l.o.g. $\vec{s}' = \{0, \dots, 0\}$ (the first probability distribution is independent of the path

afterwards). Inductively using lemma 5 one gets:

$$\begin{aligned}
& \int_0^l p(x) dx \\
& = \int_0^l p_{\vec{s}'}^0(x) dx \\
& \geq \sum_{\vec{s} \in \{0,1\}^m} \int_0^l \left(\prod_{j=1}^m \eta_{\vec{s}}^j \right) \left(\prod_{j=1}^m w_{\vec{s}}^j \right) \tau_t^m \circ p_{\vec{s}} \left(x \prod_{j=1}^m w_{\vec{s}}^j \right) \\
& \quad \int_0^l \left(\prod_{j=1}^m w_{\vec{s}}^j \right) \left(\prod_{j=1}^m \eta_{\vec{s}}^j \right) \tau_t^m \circ p_{\vec{s}}(x) \\
& = \sum_{\vec{s} \in \{0,1\}^m} \int_0^l \left(\prod_{j=1}^m \eta_{\vec{s}}^j \right) \tau_t^m \circ p_{\vec{s}}(x) \\
& \quad \int_0^l \left(\prod_{j=1}^m w_{\vec{s}}^j \right) \left(\prod_{j=1}^m \eta_{\vec{s}}^j \right) \tau_t^m \circ p_{\vec{s}}(x) \\
& \geq \sum_{\vec{s} \in G} \int_0^l \left(\prod_{j=1}^m \eta_{\vec{s}}^j \right) \tau_t^m \circ p_{\vec{s}}(x) \\
& \geq \sum_{\vec{s} \in G} \int_0^{lw} \left(\prod_{j=1}^m \eta_{\vec{s}}^j \right) \tau_t^m \circ p_{\vec{s}}(x) \\
& = P_S \int_0^{lw} q(x) dx
\end{aligned}$$

Therefore (with $P_S = 1 - \epsilon$):

$$\begin{aligned}
W^\epsilon & = kT \ln(w) \\
& \leq kT \ln \left(\max \left\{ m \left| \int_0^l p(x_1) dx_1 \geq \int_0^{lm} (1-\epsilon) q(x_2) dx_2 \forall l \right. \right\} \right) \\
& = kT \ln \left(M \left(\frac{G(\rho)}{1-\epsilon} || G(\sigma) \right) \right)
\end{aligned}$$

Which proves the main theorem. \square

II. RECOVERING THE RELATIVE MIN-ENTROPY

We now recover the result of eq. 2 which, as discussed in the main body, was given in [13, 14]. For completeness we repeat the equation also here:

$$W^\epsilon = kT \ln(2) D_0^\epsilon(\sigma || \rho_T).$$

To do this we first need the corresponding expression for the Gibb's rescaled entropies:

In the case where one starts at any state P on a system with only degenerate energy levels and ends up in the thermal state $f = (2^{-n}, \dots, 2^{-n})$, we would expect the maximal work that can be extracted by taking a risk ϵ to be $kT \ln(2) (n - H_{\max}^\epsilon(P))$, as seen in [5] for the case of Szilard engines. Indeed this is the result. More generally:

Theorem 6.

$$W^\epsilon = kT \ln(2) (H_{\max}(q) - H_{\max}^\epsilon(p))$$

where $p = G^T(\rho)$ is the Gibb's rescaled probability distribution corresponding to the initial state ρ and $q = G^T(\sigma)$ is the one corresponding to the final state σ .

For the proof of this theorem a technical lemma on the smooth max-entropy is needed.

Lemma 7. Let p be a monotonously falling probability function on $[0, \infty)$ and d_ϵ be defined through

$$\int_0^{d_\epsilon} p(x)/(1-\epsilon) dx = 1$$

Then:

$$d_\epsilon = 2^{H_{\max}^\epsilon(p)}$$

Proof. Let d_ϵ be defined as above. We need to show two things:

- i) $\exists p^\epsilon$ probability function on $[0, \infty)$ with $\|\text{supp}(p^\epsilon)\| = d_\epsilon$ and trace-distance $\delta(p, p^\epsilon) < \epsilon$.
- ii) $\|\text{supp}(p^\epsilon)\| \geq d_\epsilon \forall p^\epsilon$ monotonously decreasing probability functions on $[0, \infty)$ with $\delta(p, p^\epsilon) < \epsilon$.

Then we get that $H_{\max}^\epsilon(p) = \log_2(\min_{\delta(p, p^\epsilon) < \epsilon}(\|\text{supp}(p^\epsilon)\|)) = \log_2(d_\epsilon)$, as said in the lemma. The proof of i) goes as follows: Define $p^\epsilon(x) = p(x) \left(\int_0^{d_\epsilon} p(x) \right)^{-1}$ for $x \leq d_\epsilon$ and $p^\epsilon(x) = 0$ for $x > d_\epsilon$. This p^ϵ is therefore normalized to one, has support $[0, d_{\epsilon\text{psilon}}]$ and the following equation shows

that it is also ϵ -near to p :

$$\begin{aligned} \delta(p, p^\epsilon) &= \frac{1}{2} \left(\int_0^\infty |p^\epsilon(x) - p(x)| dx \right) \\ &= \frac{1}{2} \left(\int_0^{d_\epsilon} |p^\epsilon(x) - p(x)| dx + \int_{d_\epsilon}^\infty p(x) dx \right) \\ &= \frac{1}{2} \left(\int_0^{d_\epsilon} (p^\epsilon(x) - p(x)) dx + \int_{d_\epsilon}^\infty p(x) dx \right) \\ &= \frac{1}{2} \left(1 - \int_0^{d_\epsilon} p(x) dx + \int_{d_\epsilon}^\infty p(x) dx \right) \\ &= \int_{d_\epsilon}^\infty p(x) dx \\ &< \epsilon \end{aligned}$$

which concludes the proof of i).

For the proof of ii) assume, that: $\exists p^\epsilon$ like above, s.t.

$|\text{supp}(p^\epsilon)| \leq d_\epsilon$, then (with $\Theta = \Theta_{0,\infty}$):

$$\begin{aligned}
& \frac{1}{2} \left(\int_0^\infty |p^\epsilon(x) - p(x)| dx \right) \\
&= \frac{1}{2} \left[\int_0^{d_\epsilon} |p^\epsilon(x) - p(x)| dx + \underbrace{\int_0^\infty \underbrace{|p^\epsilon(x) - p(x)|}_{\geq \epsilon} dx}_{\geq \epsilon} \right] \\
&\geq \frac{1}{2} \left[\epsilon + \int_0^{d_\epsilon} \Theta(p(x) - p^\epsilon(x)) (p(x) - p^\epsilon(x)) dx \right] \\
&\quad + \frac{1}{2} \left[\int_0^{d_\epsilon} \Theta(p^\epsilon(x) - p(x)) (p^\epsilon(x) - p(x)) dx \right] \\
&\geq \frac{1}{2} \left[\epsilon + \int_0^{d_\epsilon} \Theta(p(x) - p^\epsilon(x)) (p(x) - p^\epsilon(x)) dx \right] \\
&\quad + \frac{1}{2} \left[\left(\int_0^\infty - \int_{d_\epsilon}^\infty - \int_0^{d_\epsilon} \Theta(p(x) - p^\epsilon(x)) \right) \right. \\
&\quad \left. (p^\epsilon(x) - p(x)) dx \right] \\
&\geq \frac{1}{2} \left[\epsilon + \underbrace{\int_0^{d_\epsilon} \Theta(p(x) - p^\epsilon(x)) (p(x) - p^\epsilon(x)) dx}_{\geq 0} \right] \\
&\quad + \frac{1}{2} \left[\underbrace{(1-1) + \int_{d_\epsilon}^\infty \Theta(p(x) - p^\epsilon(x)) (p(x) - p^\epsilon(x)) dx}_{\geq 0} \right] \\
&\quad + \frac{1}{2} \left[\underbrace{\int_{d_\epsilon}^\infty (p(x) - p^\epsilon(x)) dx}_{\geq \epsilon} \right] \\
&\geq \epsilon
\end{aligned}$$

which is a contradiction to $\delta(p, p^\epsilon) = \frac{1}{2} \left(\int_0^\infty |p^\epsilon(x) - p(x)| dx \right) < \epsilon$. \square

Now we have all we need to proof the theorem above:

Proof. let p^ϵ be a probability function with the smallest possible support such that $\delta(p, p^\epsilon) \leq \epsilon$ and define d_ϵ as in lemma 7. For $l \leq d_\epsilon$ the requirement for maximal work

extraction reads (using the lemma)

$$\begin{aligned}
\int_0^l \frac{p(x)}{1-\epsilon} dx &\geq \frac{l}{d_\epsilon} \int_0^{d_\epsilon} \frac{p(x)}{1-\epsilon} dx = \frac{l}{\|\text{supp}(q)\|} \frac{\|\text{supp}(q)\|}{\|\text{supp}(p^\epsilon)\|} \\
&= \int_0^l \frac{l \frac{\|\text{supp}(q)\|}{\|\text{supp}(p^\epsilon)\|}}{\|\text{supp}(p^\epsilon)\|} q(x) dx
\end{aligned}$$

The above is an equation in the case $l = d_\epsilon$. Which shows that the maximal w as defined in theorem 1 is given by

$$w = \frac{\|\text{supp}(q)\|}{\|\text{supp}(p^\epsilon)\|} = 2^{(H_{\max}(q) - H_{\max}^\epsilon(p))}$$

\square

Equation 2 can than be seen as a corollary of this theorem:

Corollary. Eq. 2 is a special case of the above theorem, recovered when the final state is a Gibb's state and has also the same energy eigenvalues as the initial.

$$W^\epsilon = kT \ln(2) D_0^\epsilon(\rho, \sigma^T)$$

Proof. Let p be the Gibbs-rescaled probability function corresponding to ρ and $P(j)$ the eigenvalues of ρ . Let a be the flat energy probability function corresponding to σ^T . Let $A(j) = \frac{\exp(-\frac{E(j)}{kT})}{Z}$, where $E(j)$ are the energy-eigenvalues of ρ and σ^T and Z is the corresponding partition function. This means by definition, that

$$p \left(Z \int_0^x A \left(\left\lceil \frac{y \cdot n}{A} \right\rceil \right) dy \right) = \frac{P \left(\left\lceil \frac{x \cdot n}{A} \right\rceil \right)}{A \left(\left\lceil \frac{x \cdot n}{A} \right\rceil \right) Z}$$

and likewise $a(x) = 1/Z$ (both defined for $x \in [0, Z]$). From the above theorem we get:

$$\begin{aligned}
W^\epsilon &= kT \ln(2) (H_{\max}(a) - H_{\max}^\epsilon(p)) \\
&= kT \ln(2) \left(\log_2(Z) - \log_2 \left(\inf_{\delta(p^\epsilon, p) < \epsilon} \text{supp}(p) \right) \right) \\
&= -kT \ln(2) \log_2 \left(\frac{1}{Z} \right. \\
&\quad \cdot \left. \min_{\left\{ x \mid \int_0^x P(\lceil y \cdot n \rceil) dy > 1-\epsilon \right\}} \left(Z \int_0^x A(\lceil y \cdot n \rceil) dy \right) \right) \\
&= kT \ln(2) R_0^\epsilon(P, A)
\end{aligned}$$

\square

III. ENTROPY INCREASE LAW

We now show that our thermalisation model indeed respects $\Delta S \geq \beta \Delta \langle E \rangle$, where S is the Von Neumann entropy, β the inverse temperature associated with the reservoir, and $\langle E \rangle = \sum_i \lambda_i E_i$ the expected internal energy.

Lemma 8. *In the model for thermalisation used here the following is always respected:*

$$\Delta S \geq \beta \Delta \langle E \rangle. \quad (8)$$

Proof. We firstly recall the model and define certain notation.

Recall that the thermalisation model states that when two levels, 1 and 2, are coupled to the heat bath, their ratio λ_1/λ_2 gets closer to $\exp(-\beta(E_1 - E_2))$, and the other λ 's are untouched. In our model one may concatenate several such interactions to implement any allowed multi-level interaction with the bath. It will therefore suffice to show that Eq. 8 holds for a single two-level interaction with the heat bath.

For notational convenience let the probability of being in level 1 or 2 be called $\lambda_{12} := \lambda_1 + \lambda_2$. This is then constant for the given two-level interaction with the bath. In the extreme case of the two levels interacting with the bath for an arbitrary amount of time we have $\lambda_1 := \lambda_1^T$ and $\lambda_2 := \lambda_2^T$ (T reminds us of the temperature dependence). These values must then obey the relation

$$\lambda_1^T / \lambda_2^T = \exp(-\beta(E_1 - E_2)) \quad (9)$$

We also assume without loss of generality that $E_2 \leq E_1$. This implies that $\lambda_1^T \leq 0.5\lambda_{12}$.

Now we begin to prove the statement. Firstly we simplify δS by noting that only two levels change their probabilities. We write

$$\begin{aligned} S &= - \sum_i \lambda_i \log \lambda_i \\ &= -\lambda_1 \log \lambda_1 - (\lambda_{12} - \lambda_1) \log(\lambda_{12} - \lambda_1) - \sum_{i=3}^{i_{\max}} \lambda_i \log \lambda_i \\ &\equiv S_{12} - \sum_{i=3}^{i_{\max}} \lambda_i \log \lambda_i. \end{aligned}$$

We see that in any two-level interaction

$$\Delta S = \Delta S_{12}. \quad (10)$$

It is helpful to reexpress S_{12} in terms of an actual entropy \overline{S}_{12} , so that we can use known properties of entropies to make statements about S_{12} . We let $\overline{\lambda}_1 := \lambda_1/\lambda_{12}$ and $\overline{\lambda}_2 := \lambda_2/\lambda_{12}$ such that $\overline{\lambda}_1 + \overline{\lambda}_2 = 1$. We define

$$\overline{S}_{12} := -\overline{\lambda}_1 \log \overline{\lambda}_1 - \overline{\lambda}_2 \log \overline{\lambda}_2.$$

One can then see in a few lines of algebra that

$$S_{12} = \lambda_{12} \overline{S}_{12} - \lambda_{12} \log \lambda_{12}$$

It follows that

$$\Delta S_{12} = \lambda_{12} \Delta \overline{S}_{12}. \quad (11)$$

We accordingly now want to show that $\lambda_{12} \Delta \overline{S}_{12} \geq \beta \Delta \langle E \rangle$.

We can now use the well known property of the shannon/von neumann entropy: \overline{S}_{12} is concave in $\overline{\lambda}_1 = \lambda_1/\lambda_{12}$. The function is accordingly upper bounded by any tangential line, as in Figure III. Consider the tan-

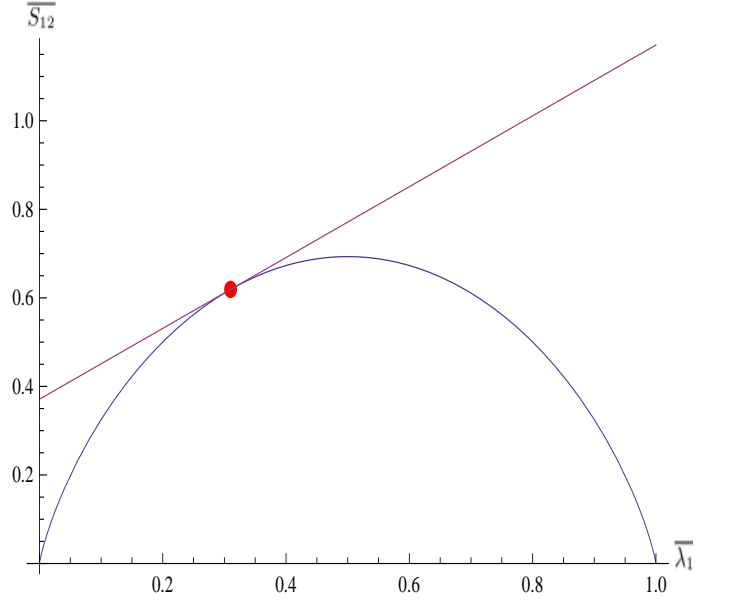


Figure 2: The entropy \overline{S}_{12} is a function of $\overline{\lambda}_1$. The red dot corresponds to the thermal state in question, i.e. $\overline{\lambda}_1 = \overline{\lambda}_1^T$. The tangential upper bound has gradient $\beta(E_2 - E_1)$.

gential line at $\lambda_1 = \lambda_1^T$. At that point it follows from a few lines that

$$\frac{d}{d\lambda_1} S_{12} |_{\lambda_1 = \lambda_1^T} = \frac{d}{d\lambda_1} \overline{S}_{12} = \beta(E_1 - E_2). \quad (12)$$

Note now that $\langle E \rangle$ may similarly to the entropy be written as

$$\begin{aligned} \langle E \rangle &= - \sum_i \lambda_i E_i \\ &\equiv \langle E \rangle_{12} + \langle E \rangle_{rest}, \end{aligned}$$

such that $\Delta \langle E \rangle = \Delta \langle E \rangle_{12} = (\Delta \lambda_1)(E_1 - E_2)$, with $\Delta \lambda_1 = \lambda_1' - \lambda_1$ the change in λ_1 . So $\langle E \rangle(\lambda_1)$ is a line with gradient given by

$$\frac{\Delta \langle E \rangle}{\Delta \lambda_1} = E_1 - E_2.$$

Similarly

$$\frac{\Delta\langle E\rangle}{\Delta\lambda_1} = \frac{1}{\lambda_{12}}(E_1 - E_2).$$

Comparing this with the gradient of the tangential line to \overline{S}_{12} in Eq. 12, we see that $\frac{1}{\lambda_{12}}\beta\langle E\rangle_{12}$ *has the same gradient* as the tangential line. We therefore only need to show that the change in the tangential line is upper bounded by the change in the entropy curve, as it is equivalent to showing that $\Delta\overline{S}_{12} \geq \frac{1}{\lambda_{12}}\beta\langle E\rangle_{12}$. This must hold for all possible initial and final values of $\overline{\lambda}_1$ and all possible values of $\overline{\lambda}_1^T$ (recall that we assumed without loss of generality that $\overline{\lambda}_1^T \geq 0.5$). These can be grouped into three cases.

1. $\overline{\lambda}_1 \leq \overline{\lambda}_1^T$. Here the tangential bound above implies that $\Delta\overline{S}_{12} \geq \frac{\beta}{\lambda_{12}}\langle E\rangle_{12} \geq 0$.
2. $\lambda_1^T \leq \overline{\lambda}_1 \leq 0.5$. Here the tangential bound implies that $0 \geq \Delta\overline{S}_{12} \geq \frac{\beta}{\lambda_{12}}\langle E\rangle_{12}$.
3. $\overline{\lambda}_1 \geq 0.5$, also after the interaction. Here the tangential bound implies that $\Delta\overline{S}_{12} \geq 0 \geq \frac{\beta}{\lambda_{12}}\langle E\rangle_{12}$.

This implies the lemma. \square